

# A ChatGPT Prompt for Writing Case-Based Multiple-Choice Questions

Yavuz Selim Kıyak<sup>1\*</sup>

<sup>1</sup> Department of Medical Education and Informatics, Gazi University Faculty of Medicine, Ankara, Turkey; [yskiyak@gazi.edu.tr](mailto:yskiyak@gazi.edu.tr), , Orcid ID: 0000-0002-5026-3234

\* Correspondence: [yskiyak@gazi.edu.tr](mailto:yskiyak@gazi.edu.tr)

Received: 8/10/2023; Accepted: 20/10/2023; Published: 23/10/23

## Abstract

The significant challenge faced by medical schools is the effortful process of writing a high quantity of high-quality case-based multiple-choice questions (MCQs) to assess the higher-order skills of medical students. The demand for a high volume of MCQs in education has led to the development of Automatic Item Generation (AIG), specifically template-based AIG, which involves creating cognitive and item models by subject matter experts to generate hundreds of MCQs at once using software. It demonstrated significant success in various languages and even being incorporated into national medical licensure exams. However, this method still heavily depends on the efforts of subject matter experts. This paper introduces a detailed ChatGPT prompt for quickly generating case-based MCQs and provides important research questions for future exploration into ChatGPT's potential in generating items, signaling the beginning of the artificial intelligence era in medical education, encouraging health professions education researchers to delve deeper into its potential.

**Keywords:** ChatGPT; automatic item generation; multiple-choice questions; artificial intelligence; medical education

---

## Introduction

Medical schools use various written assessment methods to evaluate higher-order skills of medical students, particularly clinical reasoning. These methods encompass case-based multiple-choice questions (MCQs), key-feature questions, script concordance tests, and others (1). Among these, case-based MCQ stands out as both highly effective and commonly utilized (2). However, the process of writing high-quality items (questions) presents a significant challenge on the limited resources of medical schools.

More specifically, writing a single context-rich MCQ designed to assess higher-order skills demands a significant investment of a medical teacher's time, often extending into hours (3). While it requires an effortful process, the demand for a substantial quantity of MCQs within medical education significantly increases the level of challenge. Even developing a question bank solely for progress testing necessitates incorporating thousands of questions (4). Hence, the significant challenge faced by medical schools lies in writing a high quantity of high-quality MCQs.

### *Existing Solutions*

High volume of MCQ need in education led researchers to develop new methods to generate high number of items at once by using software. Automatic Item Generation (AIG) has been developed for this purpose as "a process of using models to generate items using computer technology" (5). Template-based AIG is based on creating a cognitive model, which reveals subject matter experts' thinking process when they faced with a clinical problem. Then, subject matter experts create an item model based on the cognitive model on a disease/problem. In the item model, a template MCQ is built, and the words

that can be manipulated in the item are determined. Finally, a software generates items by assigning the appropriate words to the appropriate parts in the items considering the specifications provided by subject matter experts (6). An item model is able to result in hundreds of MCQs, therefore template-based AIG is more feasible than traditional way of item writing.

Template-based AIG has shown a great success in the last ten years. Successful implementations have been shown in the literature in various languages, such as English, French, Chinese, Spanish, Korean, and Turkish (6,7). Furthermore, automatically generated questions have been used in national medical licensure exam in Canada (8).

Despite the superiority of template-based AIG over traditional item writing, it still depends on human efforts as it requires subject matter experts to develop cognitive models and item models. Although artificial intelligence (AI) is less effortful for test developers to generate items, researchers still preferred template-based AIG over the use of AI for this purpose because AI provided unsatisfying MCQs compared to template-based methods (6).

#### *AI for Generating MCQs*

Non-template based methods in AIG relies on Natural Language Processing (NLP), which has been used but did not produce satisfying MCQs in the past (6). However, it seems NLP-based AI solutions like generative pretrained transformer (GPT), which is a large language model (LLM), is about to change this. It is because LLMs, such as GPT-3.5, achieved a success in “hacking our languages” (9).

ChatGPT, a variant of GPT-3.5, was publicly introduced in November 2022. It passed one million user mark in only five days (10). It affected health professions education as well (11–14). As a result, some medical education researchers suggested the release date of ChatGPT as a crucial delineation between the pre-ChatGPT era and the post-ChatGPT era (15). This success was probably due to the fact that GPT-3 possesses a data corpus that is significantly larger, being ten times the size of prior models (16), and due to ChatGPT’s free and public release (15).

Our personal experiences show that generating high quality MCQs using ChatGPT depends on the quality of prompts. Therefore, we developed a prompt by revising the prompt developed by Esh Tatla for medical students (17). Even if the original prompt was useful, it required us to make revisions to generate MCQs that align more closely with the principles of constructing multiple-choice items in medical education. The prompt, achieved through continuous testing and refinement, is presented in Table 1. It also includes a part to provide explanations of the answers that may help test developers and medical students to better understand.

To use this prompt, the user needs only to fill two parts presented using capital letters:

1. [PLEASE INSERT A TOPIC]
2. [PLEASE INSERT A DIFFICULTY LEVEL (E.G. EASY, DIFFICULT)]

In the first part, the user inserts a topic. The topic can also be considered as a learning objective of a training in a medical school. The second part is to determine the difficulty of the MCQ.

Table 2 shows an MCQ generated by using the prompt with “managing essential hypertension in a primary care setting” as the topic, and “Easy” as the difficulty level, which is accessible from <https://chat.openai.com/share/4342df6f-7cbe-43df-98dc-b110b53983e3>. (Free Research Preview, September 25 Version).

**Table 1.** A prompt for generating MCQs in medical education.

You are developing a question bank for medical exams focusing on the topic of [PLEASE INSERT A TOPIC]. Please generate a high-quality single best answer multiple-choice question. Follow the principles of constructing multiple-choice items in medical education. Generate the questions using the following framework:

Case (write as a single narrative paragraph without providing each part separately):

Patient details (gender/age)

Presenting complaint

Relevant clinical history

Physical examination findings

Diagnostic test results (optional)

Question stem: [Insert relevant information from the above sections without compromising the answer]

Acceptable question style: Ask for the BEST answer, NOT one that is TRUE/FALSE.

Answer options:

[Insert plausible answer option]

Explanation:

- Identify and explain the correct answer.
- Explain why this is the most appropriate answer based on evidence-based guidelines or expert consensus.
- Briefly explain why the other answer options are less correct or incorrect.

Difficulty level: [PLEASE INSERT A DIFFICULTY LEVEL (E.G. EASY, DIFFICULT)]

**Table 2.** An MCQ generated by using the prompt in ChatGPT.**Case:**

Mr. Johnson, a 55-year-old male, presents to the primary care clinic with a chief complaint of recurrent headaches. He reports a sedentary lifestyle and a family history of hypertension. His blood pressure readings during previous routine check-ups have consistently been elevated. Mr. Johnson denies any chest pain, shortness of breath, or visual disturbances. His past medical history is significant for obesity and a sedentary occupation. On physical examination, his blood pressure is 150/92 mmHg, and his BMI is 31. Laboratory results show normal renal function and lipid profile.

**Question stem:**

Given Mr. Johnson's presentation with recurrent headaches, a family history of hypertension, and consistently elevated blood pressure readings, what is the BEST initial management strategy in a primary care setting?

**Answer options:**

- A. Initiate lifestyle modifications, including dietary changes and increased physical activity.
- B. Start antihypertensive medication immediately to lower blood pressure.
- C. Order a 24-hour ambulatory blood pressure monitoring for further assessment.
- D. Recommend immediate admission to the hospital for comprehensive evaluation.
- E. Suggest over-the-counter analgesics for headache relief and reassess in six months.

**Explanation:**

The correct answer is A. Initiate lifestyle modifications, including dietary changes and increased physical activity.

**Explanation:**

Hypertension management guidelines emphasize lifestyle modifications as the initial approach in patients with uncomplicated essential hypertension. For individuals like Mr. Johnson, who presents with elevated blood pressure, obesity, and a sedentary lifestyle, initiating lifestyle changes is crucial. These changes should include adopting a heart-healthy diet (such as the DASH diet), increasing physical activity, and weight reduction. Lifestyle modifications have been shown to significantly contribute to blood pressure control and may obviate the need for immediate pharmacologic intervention.

Option B (Start antihypertensive medication immediately) is not the best initial approach unless there are compelling indications for immediate treatment, which are not evident in Mr. Johnson's case. Option C (Order a 24-hour ambulatory blood pressure monitoring) may be considered in certain situations but is not the primary step in the management of essential hypertension. Option D (Recommend immediate admission to the hospital) is not warranted in the absence of hypertensive emergencies. Option E (Suggest over-the-counter analgesics) does not address the underlying hypertension and is not the primary management strategy for this patient.

As a physician with a PhD in medical education, it seems the MCQ above is both plausible and well-constructed. While a thorough analysis of the item's quality could be carried out, it is initially intriguing to generate such an MCQ within seconds by simply inputting a prompt and pressing a button, which is an extraordinary success. In the past, writing such MCQs required extensive time of subject matter experts. However, in today,

with an internet connection, an average computer, and the detailed prompt, we can efficiently generate MCQs in a few seconds. This is why some researchers name this period began in November 2022 as post-ChatGPT era.

### Recommendations for Future Research

Automatic MCQ generation using ChatGPT comes up with many research questions that need to be answered. Some of them are:

- Are the MCQs considered acceptable through evaluation by subject matter experts in terms of scientific and clinical knowledge?
- Is the prompt capable of generating MCQs that meet psychometric standards?
- How does the quality MCQs generated by ChatGPT compared to those created by human experts?
- How do the cognitive levels of questions generated by ChatGPT align with the intended learning objectives and Bloom's taxonomy?
- Do the intended difficulty levels of the MCQs align with their actual difficulty in real assessment settings?
- Can the prompt be effectively used to generate MCQs in various health professions education settings?
- Do the MCQs need editing by humans before incorporating them into exams?
- Do the explanations serve as effective feedback for medical students in formative assessments?

### Conclusions

- ChatGPT provides an exciting opportunity for generating case-based MCQs in a few seconds.
- Researchers should conduct studies to evaluate the quality of MCQs generated by using ChatGPT.

**Funding:** There has been no funding.

**Declaration of conflict of interest:** The author declares that there is no conflict of interest.

### References

1. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine*. 2019;94(6): 902–912. <https://doi.org/10.1097/ACM.0000000000002618>.
2. Pugh D, De Champlain A, Touchie C. Plus ça change, plus c'est pareil: Making a continued case for the use of MCQs in medical education. *Medical Teacher*. 2019;41(5): 569–577. <https://doi.org/10.1080/0142159X.2018.1505035>.
3. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education*. 2004;38(9): 974–979. <https://doi.org/10.1111/j.1365-2929.2004.01916.x>.
4. Wrigley W, Van Der Vleuten CP, Freeman A, Muijtjens A. A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71. *Medical Teacher*. 2012;34(9): 683–697. <https://doi.org/10.3109/0142159X.2012.704437>.
5. Gierl MJ, Lai H, Turner SR. Using automatic item generation to create multiple-choice test items. *Medical Education*. 2012;46(8): 757–765. <https://doi.org/10.1111/j.1365-2923.2012.04289.x>.
6. Gierl MJ, Lai H, Tanygin V. *Advanced Methods in Automatic Item Generation..* 1st ed. Routledge; 2021. <https://doi.org/10.4324/9781003025634> [Accessed 1st August 2022].

7. Kiyak YS, Budakoğlu İ, Coşkun Ö, Koyun E. The First Automatic Item Generation in Turkish for Assessment of Clinical Reasoning in Medical Education. *Tıp Eğitimi Dünyası*. 2023;22(66): 72–90. <https://doi.org/10.25282/ted.1225814>.
8. Gierl MJ, Lai H, Pugh D, Touchie C, Boulais AP, De Champlain A. Evaluating the Psychometric Characteristics of Generated Multiple-Choice Test Items. *Applied Measurement in Education*. 2016;29(3): 196–210. <https://doi.org/10.1080/08957347.2016.1171768>.
9. Harari YN. Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. *The Economist*. 2023; <https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation> [Accessed 7th October 2023].
10. Buchholz K. *Infographic: ChatGPT Sprints to One Million Users*. Statista Infographics. <https://www.statista.com/chart/29174/time-to-one-million-users> [Accessed 28th April 2023].
11. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students. *European Journal of Dental Education*. 2023; eje.12937. <https://doi.org/10.1111/eje.12937>.
12. Carrasco JP, García E, Sánchez DA, Porter E, De La Puente L, Navarro J, et al. ¿Es capaz “ChatGPT” de aprobar el examen MIR de 2022? Implicaciones de la inteligencia artificial en la educación médica en España. *Revista Española de Educación Médica*. 2023;4(1). <https://doi.org/10.6018/edumed.556511>.
13. Choi EPH, Lee JJ, Ho MH, Kwok JYY, Lok KYW. Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. *Nurse Education Today*. 2023;125: 105796. <https://doi.org/10.1016/j.nedt.2023.105796>.
14. Wang YM, Shen HW, Chen TJ. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. *Journal of the Chinese Medical Association*. 2023;86(7): 653–658. <https://doi.org/10.1097/JCMA.0000000000000942>.
15. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No.158. *Medical Teacher*. 2023; 1–11. <https://doi.org/10.1080/0142159X.2023.2186203>.
16. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language Models are Few-Shot Learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds.) *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; 2020. p. 1877–1901. [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf)
17. Tatla E. *5 Essential AI (ChatGPT) Prompts Every Medical Student and Doctor Should be Using to 10x their...* Medium. 2023. <https://medium.com/@eshtatla/5-essential-ai-chatgpt-prompts-every-medical-student-and-doctor-should-be-using-to-10x-their-de3f97d3802a> [Accessed 18th September 2023].



© 2023 Universidad de Murcia. Enviado para su publicación en acceso abierto bajo los términos y condiciones de la licencia Creative Commons Reconocimiento-NoComercial-Sin Obra Derivada 4.0 España (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).